SYNOPSIS

A Hybrid Machine Learning Approach for Detecting Intentional and Unintentional Insider Threats with Mitigation through Behavioral Biometrics and User Profiling Mechanism

Submitted by

S. Asha

(20PHCSF005)

Under the Guidance of

Dr.D. Shanmugapriya M.Sc., M.Phil., Ph.D., SET
Assistant Professor and Head,
Department of Information Technology
Avinashilingam
Institute For Home Science and Higher Education for Women,
Coimbatore – 641 043

Submitted to

AVINASHILINGAM
Institute For Home Science and Higher Education for Women
Coimbatore – 641 043

January 2025

A HYBRID MACHINE LEARNING APPROACH FOR DETECTING INTENTIONAL AND UNINTENTIONAL INSIDER THREATS WITH MITIGATION THROUGH BEHAVIORAL BIOMETRICS AND USER PROFILING MECHANISM

ABSTRACT

Insider threat is a potential threat to an organization that results in financial and reputation losses while exposing sensitive information. Past research extensively focused on external threats, and overlooked on both intentional and unintentional insider threats. Several researchers majorly focused on detecting such insider activities but fail to mitigate both intentional and unintentional insider threats. Few challenges such as mishandling imbalanced dataset and fail to incorporate novel feature engineering techniques, limited mitigation strategies are encountered. A methodology comprising of three phases is proposed. In Preprocessing and Insider Detection (P&ID) phase, log data is preprocessed using data integration, encoding and tuned the nearmiss-2 sampling technique to diminish the class imbalance problem. The preprocessed data is applied in B-SVM technique to detect and classify users into genuine, intentional insiders, and unintentional insiders. P&ID achieved 99.15% detection accuracy, with a low misclassification rate of 0.85% for detecting both intentional and unintentional insider threats.

The unintentional insiders are authenticated in Unintentional Insider Mitigation (UIM) phase, where Clonal Kernel Principal Component Analysis (CKPCA) along DBN is proposed to improve feature representation for keystroke authentication. UIM phase achieved 99.84% authentication accuracy and a 0.15% Equal Error Rate (EER) of 0.15% for authenticating the unintentional insiders and classify them into genuine and intentional insiders. In Intentional Insider Mitigation, the detected intentional insiders from P&ID and UIM phases are profiled into denylist and genuine users are profiled into allowlist using Decision tree.

The proposed methodology is validated using CERT Insider Threat Dataset and the CIC Darknet Dataset. P&ID detected 8 intentional and one unintentional insider among 250,078 daily logs using CERT Dataset. P&ID is validated with darknet dataset, detected 4783 intentional-Darknet users and 68 unintentional-Darknet users where (VPN: 42) (Tor: 21) (NonVPN: 5) among 134,305 daily activities. UIM mitigated one unintentional insider as an intentional insider using CERT log activities. UIM mitigated 68 unintentional-Darknet users as 64 Intentional-Darknet and 4 benign users using darknet dataset. IIM profiled 56 genuine users in allowlist and 9 intentional insiders in denylist using CERT dataset. Using CIC

Darknet dataset, the IIM profiled 5063 benign users in allowlist and 4847 Intentional-Darknet users in denylist.

INTRODUCTION

Insider threat is considered for its passive threat vector [1] in cybersecurity, which is unnoticeable and vulnerable to security systems. Gaining impulsive control toward sensitive information using legitimate authentication by a fellow employee, resulting in data exposure is a common trait of insider threat [2]. However, the nature of attaining targeted credentials is the only way to differentiate intentional and unintentional insiders. i.e., an intentional insider performs perceptive malicious activities [3], whereas an unintentional insider performs susceptible careless activities. As a consequence, commercial damage regulates brand popularity.

Forbes declared that insiders are accompanied by hacker group called Lapsus and confidently perform prognosticating, social engineering, snooping, and corporate malevolence, to augment the organizational attacks by 2023 [4]. In 2022, insider threat resulted in immense data breaches including Optus telecommunications, WhatsApp, Medibank, Rockstar, Uber, Twitter, Nelnet Servicing, Zoetop e-commerce, and Revolut financial technology [5]. According to CrowdStrike, a security-based software company [6], 49% of intentional data breaches are for financial gain, 18% of targeted outsiders, and 1% of momentum data breaches. As per 2022 Data Breach Investigation Report [7], individual-oriented data breach is 82% where 80% of them are outsider threat vectors. However, 76% are insider driven security breaches and 96% for self interest.

The success of the threat rate for such insider threat events is higher by misusing their legitimate trustworthiness in a security system [8]. The significance of minor insider threats is higher in case of insignificant detection [9]. Recognizing such insider threats is difficult due to the dynamic nature of attack origin.

Machine Learning (ML) applications have been widely used in an eminent domain of insider threats [10]. ML models are sensitive to minimal abnormal behavior, often treating it as an outlier and excluding it from classification [11]. Therefore, analyzing imbalanced data using ML algorithms is a challenging one [12], and it fails to achieve a nominal misclassification rate for malicious insider threat detection [13]. i.e., a genuine user activity is wrongly recognized as malicious behavior [14]. The consequence of such user behavior misinterpretation is more vulnerable in a live environment.

It requires a novel algorithm for detecting both intentional and unintentional insider threats. However, for insider threat mitigation, existing research recommended a user authentication mechanism using keystroke dynamics where an optimization technique is required for enhanced performance.

In this research, a novel methodology for recognizing and mitigating both intentional and unintentional insider threats is proposed to subdue the above-mentioned challenges by introducing a three-phase methodology: 1) Pre-processing & Insider Detection (P&ID), which preprocesses log data and uses a hybrid Support Vector Machine (SVM) and Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) algorithms to detect and classify users into intentional and unintentional insider based on abnormal behavior; 2) Unintentional-insider Mitigation (UIM), which authenticates unintentional insiders using behavioral biometric analysis via keystroke dynamics and enhanced the performance using Clonal-Kernel Principal Component Analysis (CKPCA); and 3) Intentional-insider Mitigation (IIM), which profiles users to differentiate between benign and malicious behavior. The methodology achieves high accuracy in detecting and mitigating insider threats, as validated on real-world datasets. The major contribution of this research is explained below:

• Nearmiss-2-based sampling techniques have been scrutinized to solve the class imbalance problem while recognizing insider threats in the P&ID phase.

• B-SVM algorithm has been employed to detect and classify both intentional and unintentional insiders in the P&ID phase with a minimal misclassification rate.

• CKPCA-DBN has been proposed in UIM phase for enhancing keystroke authentication using deep learning techniques against unintentional insiders.

• A profiling mechanism is integrated into the IIM phase for mitigating genuine users into allowlist and intentional insiders into denylist.

PROBLEM STATEMENT

To develop an approach that detects and mitigates both intentional and unintentional insider threats.

OBJECTIVES

Primary Objective

To detect and classify both intentional and unintentional insider threats.

Secondary Objectives

The major secondary objectives formulated are as follows.

- To detect and classify both intentional and unintentional insider threats by effectively addressing the challenges of class imbalance and reducing misclassification rate.
- To mitigate unintentional insiders by enhancing detection accuracy and minimized error rates in user authentication system.
- To mitigate intentional insiders through user profiling mechanism techniques.

METHODOLOGY

The present research introduces a novel methodology to combat the pressing issue of detecting and mitigating both intentional and unintentional insider threats. The methodology consists of three phases:

Phase 1: Pre-processing & Insider Detection (P&ID)

Figure 1 depicts the flow diagram of the P&ID phase. The P&ID consist of two sub-phases namely preprocessing and insider detection. Data integration, data encoding and sampling is done in preprocessing to handle class imbalance problem. Insider detection is done by B-SVM algorithm to detect and classify intentional, unintentional insiders and genuine users.

The scalability of the proposed methodology is rigorously examined using the CERT insider threat dataset [16] to analyse the publicly available network activities to detect intentional and unintentional insider detection. The adaptability of the proposed methodology is validated by analysing the hidden and anonymous network activities in VPN and Tor using the CIC Darknet dataset [17] to detect intentional and unintentional abnormal activities. The steps involved in the P&ID phase is briefly described below.

- Step 1: Combine the log data from dataset in data integration.
- Step 2: Transform the categorical values into numerical format using the data encoding algorithm.

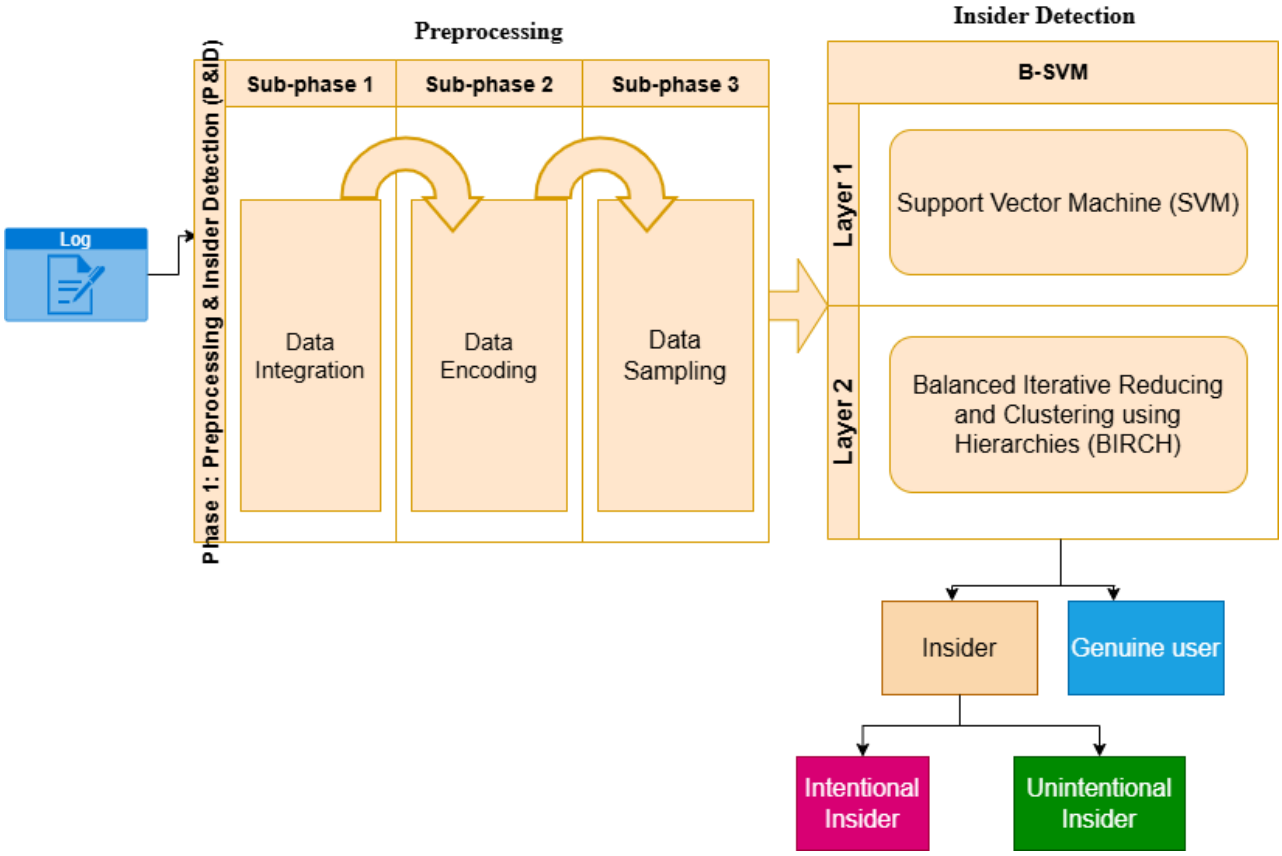

Figure 1: Methodology Overview of P&ID

- Step 3: Nearmiss-2 sampling technique is applied to combat class imbalance problem.
- Step 4: The balanced data is analyzed using B-SVM algorithm to detect and classify genuine, intentional, and unintentional insiders. In layer 1, the SVM algorithm is used for intentional insider detection that recognizes genuine and intentional insiders based on user activity. In layer 2, the BIRCH algorithm is used for unintentional insider detection that analyses falsely detected intentional insider activities from layer 1 and classifies them into genuine, intentional, and unintentional insiders.

Phase 2: Unintentional-insider Mitigation (UIM)

Detected unintentional insiders undergo authentication in the UIM using combination of Clonal-Kernel Principal-Component-Analysis (CKPCA) and Deep Belief Network (DBN) for mitigating unintentional insiders. The flow diagram of the UIM phase is depicted in Figure 2.

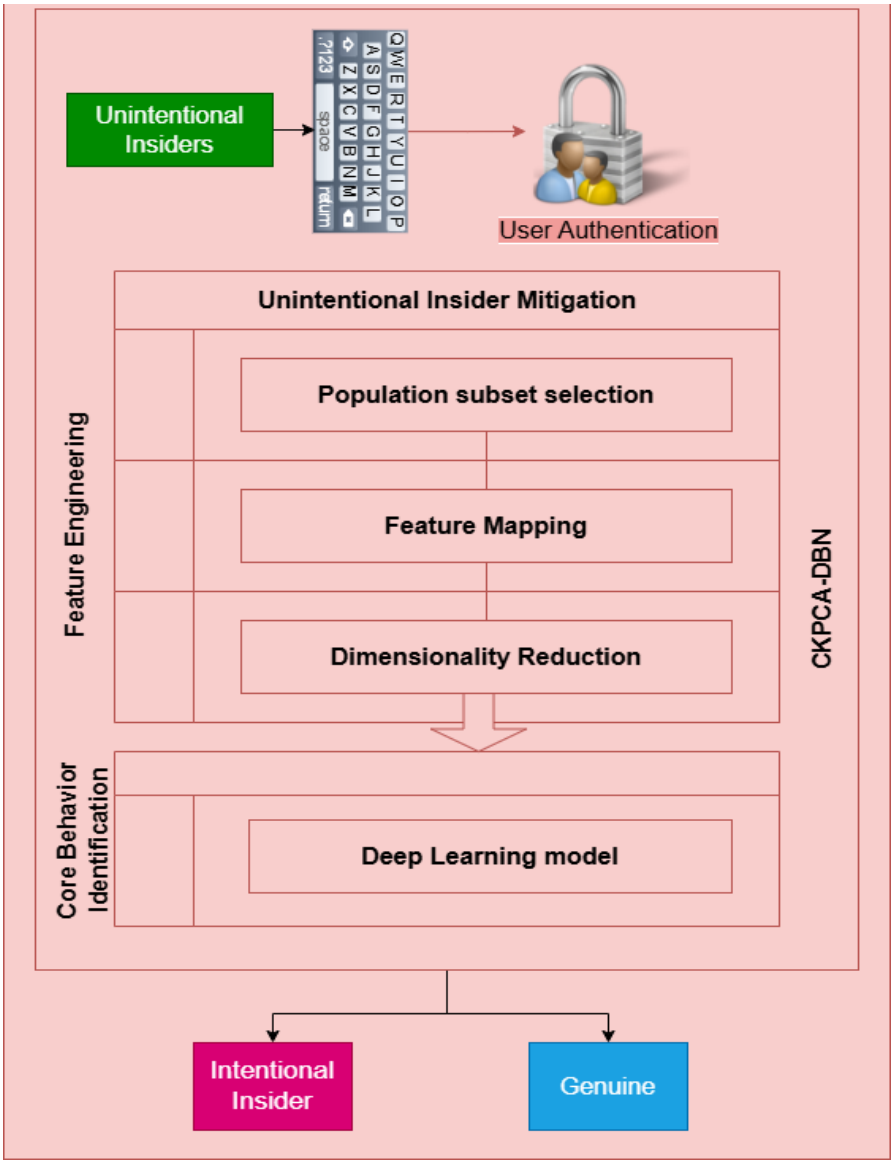

Figure 2: Methodology Overview of UIM

Figure 2 shows that the proposed methodology for phase II contains two layers namely feature engineering and core behavior identification.

- In the first layer, feature engineering technique namely CKPCA is proposed comprising of population subset selection, feature mapping and dimensionality reduction to effectively obtain intricate pattern.
- In second layer, the DBN algorithm is trained to identify the core behavior of unintentional insider using intricate pattern and authenticate the user into genuine and intentional insider.

The performance of UIM is evaluated using two datasets such as: benchmark dataset from Carnegie Mellon University [15], and real dataset.

The steps involved in UIM are described below.

- Step 1: In CKPCA, the user information containing keystroke features are first processed using Clonal Selection Algorithm for population subset selection to identify the best population among them.
- Step 2: Feature is mapped using kernel mean embedding to identify in-depth variations in selected keystroke population and map them into high dimensional structure.
- Step 3: High Dimension features are reduced using PCA to find the abstract of engineered features that helps to increase model accuracy.
- Step 4: DBN is applied to define behavioral core of each user with an aim to detect anomalous behavior. It classifies the unintentional insider into genuine and imposter (intentional insider) based on user authentication.

Phase 3: Intentional-insider Mitigation (IIM)

Mitigation of unintentional insiders is accomplished in previous phase. However, detected intentional insiders from P&ID and UIM phase undergo mitigation in IIM which are discussed below. Users are profiled into allowlists and denylists based on their behavior patterns. This phase effectively distinguishes between benign and malicious users, mitigating intentional insider threats using IIM phase.

Figure 3 depicts the flow diagram of the IIM phase. The IIM consist of three layers namely data preprocessing, model training and evaluation, and user profiling. In data preprocessing, label encoding and train test split is done. In model training and evaluation, a Decision Tree classifier is trained and evaluated to predict the user risk. In user profiling, the trained Decision Tree classifier is used for profiling the users into Allowlist if user possesses low risk, profiled into Denylist if user possesses high risk.

The steps involved in IIM phase is described below.

- Step 1: The detected genuine users in P&ID and UIM phase are listed and profiled in allowlist due to their authentic behavior pattern.
- Step 2: The detected intentional insiders in P&ID and UIM phase are recognized and profiled in denylist due to their malicious activities.

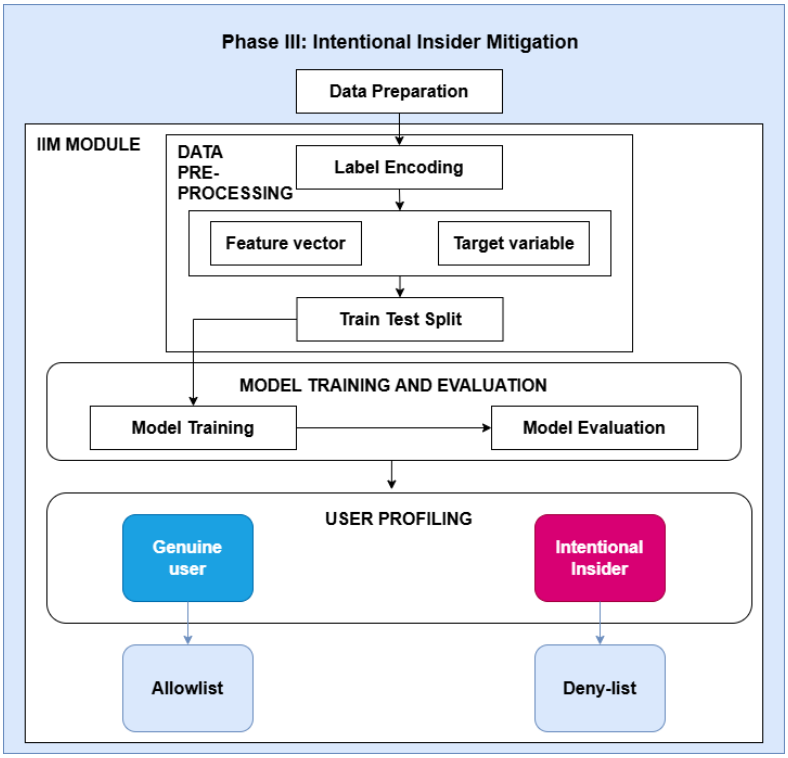

Figure 3: Methodology Overview of IIM

ELABORATIVE RESULTS

The working of the three phases such as P&ID, UIM and IIM is evaluated using diverse performance metrics. Table 1 shows the performance metrics to evaluate the performance of the proposed methodology incorporating ML and DL algorithms in Phase I, II, and III.

Table 1: Performance metrics used

Formula	Description
Precision= TP/ (TP+ FP)	Rate of threat instances considered as insider.
Recall = TP / (TP+ FN)	Recall or sensitivity, is a rate of threat instances correctly classified as insider
F-score = 2 * (precision * recall) / (precision + recall)	Also named F-measure and it is defined as the harmonic mean between recall and precision.
Accuracy = TP+TN/(TP+TN+FP+FN)	Rate of correctly detected positives and negative instances.
Misclassification rate (or false alarm rate) =	Rate of imprecisely recognized both genuine and malicious activities.
FPR = FP/(TN+FP)	Rate of incorrectly detected genuine activities.
FNR = FN/(TP+FN)	Rate of incorrectly recognized malicious activities.
TNR=TN/(TN+FP)	Rate of correctly detected genuine activities
TPR=TP/(TP+FN)	Rate of correctly recognized malicious activities

False Acceptance Rate (FAR) = FP/(FP+TN)	FAR is defined as a rate of imprecise detection of unauthorized user.
False Rejection Rate (FRR) = FN/(FN+TP)	FRR denotes the probability of inaccurately recognizing authorized user.
Equal Error Rate (EER)= (FAR+FRR)/2	EER defines the intersection of FAR and FRR.
Accuracy = (TN+TP)/(TN+TP+FN+FP)	For every recognition, the total number of precise detection of authorized and unauthorized users is considered as accuracy.

DATASET

Two datasets are used to validate the proposed methodology. Insider threat dataset from CERT division at Carnegie Mellon University in collaboration with ExactData, LLC [16] contains log activities of 4000 users in 516 days. The log activities are signing in, website perusing, document acquiring, email addressing, analyzing psychographs, and lightweight directory access protocol. The CIC-Darknet2020 dataset [17] is an amalgamation of two public datasets from the University of New Brunswick. It combines the ISCXTor2016 and ISCXVPN2016 datasets, which consist traffic details of Tor, non-Tor, VPN, and non-VPN.

EXPERIMENTAL RESULTS

The phase-wise results are demonstrated in this section.

(i) Elaborative results of P&ID

The experimental result of data preprocessing in the P&ID phase using both the CERT insider dataset and the CIC darknet dataset is mentioned in the following.

i) Data Integration

Table 3 shows the result of the data integration method using the simple feature concatenation technique. The dataset used is a comprehensive collection of network log activities that includes traffic data from Virtual Private Network (VPN) sources, The Onion Router (Tor), and non-VPN connections. Since, it is combined no integration is required.

Table 3. Data integration using simple feature concatenation technique in CERT dataset**.**

Feature name	**Description**
Insider threat	The activity classified as malicious activity or not
Vector	The origin of activity
Date	Date of each activity performed
User	Identifier of each user
Pc	Identifier of each computer to perform specific activity
Activity	Information about specific action

ii) Data Encoding

The results of data encoding using categorical encoding in the CERT and CIC Darnet dataset are mentioned in Table 4.

Table 4. Result of data encoding technique using categorical encoding.

Dataset	Attribute	Before encoding		After encoding	
		Datatype	Value	Datatype	Value
CERT Insider Dataset	Insiderthreat	Categorical	Genuine, malicious	Integer	0,1
	Source	Categorical	Logon, device, http	Integer	0,1,2
	Timestamp	Ordinal	07-01-2010 02:23:00	Integer	1280707200
	User id	Categorical	CCH0959	Integer	4
	Pc id	Categorical	PC-0588	Integer	128
	Activity	Categorical	http://linkedin.com/jobs/displayhome.html	Integer	750
CIC Darknet Dataset	Flow ID	Categorical	10.135.135.11-216.58.220.99-57158-443-6	Integer	20738
	Src IP	Categorical	10.135.135.11	Integer	13
	Dst IP	Categorical	216.58.220.99	Integer	2973
	Timestamp	Ordinal	24/07/2015 04:09:48 PM	Integer	4532
	Label	Categorical	Non-Tor, NonVPN, VPN, Tor	Integer	0, 1, 2, 3
	Label.1	Categorical	P2P, Browsing, Audio-Streaming, Chat, File-Transfer, Video-Streaming, Email, VOIP	Integer	0, 1, 2, 3, 4, 5, 6, 7

From the table 4, it is observed that user, PC, and activity in CERT and Flow ID, Src IP, Dst IP, Timestamp, Label, and Label.1 in CIC darknet dataset is converted into a numerical representation. Timestamp is converted into several milliseconds.

iii) Nearmiss-2 based sampling

After data encoding, sampling is performed using the tuned Nearmiss-2 using both the CERT dataset and the CIC darknet dataset and results are displayed in table 5.

Table 5: Data description of CERT dataset using Nearmiss-2 technique.

Dataset	**Training Set**	**Before Sampling**	**After Sampling**
CERT Insider Dataset	Genuine behavior Majority instance	(0, 2,50,078)	(0, 280)
	Malicious behavior minority instance	(1, 280)	(1, 280)
CIC Darknet Dataset	Genuine behavior instance (Majority Class: Non-Tor)	(0, 110442)	(0, 1392)
	Genuine behavior instance (Minority Class: NonVPN)	(1, 23863)	(1, 23863)
	Malicious behavior instance (Majority class: VPN)	(2, 22919)	(2, 1392)
	Malicious behavior instance (Minority class: Tor)	(3, 1392)	(3, 22919)

From the table 5, it is observed that nearimiss-2 sampling has reduced the existing 2,50,078 genuine instances to 280 instances using CERT dataset. In CIC Darknet dataset, Nearmiss-2 sampling has reduced the existing 110442 Non-Tor instances to 1392 aligned with VPN instances. The Tor instances were increased to match the NonVPN instances.

iv) B-SVM for classification

The following Table 6 gives the performance evaluation of B-SVM in the P&ID phase with SVM and BIRCH using CERT and Darknet datasets.

Table 6: Performance evaluation of P&ID phase

Dataset	Algorithm	Precision (%)	Recall (%)	F-score (%)	Accuracy (%)	Misclassification rate (%)	FPR (%)	FNR (%)	TNR (%)	TPR (%)
CERT dataset	SVM	88	96.03	91.84	91.46	8.53	12	4.36	88	95.63
	BIRCH	100	33.34	50	66.67	33.33	0	40	100	60
	B-SVM	**100**	**98.01**	**98.99**	**99.01**	**0.99**	**0**	**1.94**	**100**	**98.05**
Darknet dataset	SVM	95.7	95.42	95.39	95.42	4.58	1.51	4.79	98.48	95.2
	BIRCH	67.15	54.25	49.33	54. 25	45.75	15.24	45.74	84.75	54.25
	B-SVM	**95.90**	**95.69**	**95.59**	**95.69**	**4.31**	**1.44**	**4.52**	**98.55**	**95.47**

From the above table 6, the following observations are noted:

- SVM using the CERT dataset showed a relatively high misclassification rate of 8.53%, with 12% FPR and 4.36% FNR.
- Using the CERT dataset, BIRCH struggled with recall and F-score and obtained a high misclassification rate.
- B-SVM using the CERT dataset outperformed both SVM and BIRCH with the highest precision, recall, F-score, and accuracy with minimal misclassification rate.
- In the darknet dataset, the performance of B-SVM surpassed SVM and BIRCH by obtaining the lowest misclassification rate.

The outcome of the P&ID phase is 8 intentional and 1 unintentional insider activity among 250,078 daily log entries is detected in CERT Dataset. On the Darknet dataset, the activities of 4,783 intentional-Darknet users and 68 unintentional-Darknet users such as VPN (42 users), Tor (21 users), and NonVPN (5 users) among 134,305 daily activities are detected.

(ii) Elaborative results of UIM

The elaborative results of the UIM phase are depicted in Table 7.

Table 7: Performance Evaluation Results of UIM phase

Dataset	DL model	FAR (%)	FRR (%)	EER (%)	Accuracy (%)
CMU Keystroke Dataset	**CKPCA-DBN**	**0.002**	**8.49**	**0.152**	**99.85**
	CSA-DBN	0.156	16.74	0.449	99.55
	CSAPCA-DBN	0.17	16.20	0.45	99.54
	PCA-DBN	0.2109	29.41	0.73	99.27
	DBN	0.15	14.81	0.41	99.58
	CKPCA-CNN	**0.52**	**9.53**	**0.66**	**99.34**
	CSA-CNN	0.421	12.25	6.33	93.66
	CSAPCA-CNN	0.97	19.50	1.201	98.79
	PCA-CNN	1.148	20.44	1.37	98.62
	CNN	0.64	11.61	0.789	99.21
Collected Dataset	**CKPCA-DBN**	**0.0011**	**2.47**	**0.081**	**99.91**
	CSA-DBN	0.0078	9.996	3.24	96.75
	CSAPCA-DBN	0.0279	1.838	0.93	99.06
	PCA-DBN	0.0849	8.92	0.37	99.62
	DBN	2.62	66.35	11.27	88.72

	CKPCA-CNN	**0.27**	**2.34**	**0.34**	**99.65**
	CSA-CNN	3.23	0	3.23	96.76
	CSAPCA-CNN	0.57	5.42	2.99	97
	PCA-CNN	1.49	8.35	1.65	98.34
	CNN	3.235	0	3.23	96.76

From table 7, the followings are observed:

- Unintentional insider authentication using DBN shows poor performance, while CNN achieved relatively better results but enhancement is still required.
- CSA-DBN obtained high FRR of 9.97%, EER of 3.24%, accuracy of 96.75%.
- The performance of CKPCA and DBN demonstrates better performance by achieving higher accuracy compared to PCA and DBN.
- CKPCA-CNN outperformed other CNN configurations but fell slightly short of CKPCA-DBN in terms of accuracy and error rates.

Meanwhile, the following observations are made using CMU keystroke dataset:

- CKPCA-DBN achieves the best performance by obtaining FAR of 0.02%, FRR of 0.84%, EER of 0.15%, accuracy of 99.84%.
- The performance of CKPCA-CNN surpasses CSAPCA-CNN in terms of increased FAR and FRR.
- CKPCA-CNN demonstrates robust performance by achieving 99.34% accuracy but falls short of CKPCA+DBN due to slightly higher FAR and EER of 0.52%, and 0.65%.

Additionally, the UIM phase mitigated one unintentional insider as an intentional insider in the CERT dataset. In the Darknet dataset, the UIM phase mitigated 68 unintentional-Darknet users as 64 Intentional-Darknet and 4 benign users.

(iii) Elaborative results of IIM

IIM phase profiled 57 genuine users in the Allowlist and eight intentional insiders in the Denylist using the CERT dataset. Using the CIC Darknet dataset, the IIM phase profiled 5063 benign users in the Allowlist and 4847 intentional darknet users in the Denylist. The validation process includes a train-test split of 80%-20%. The methodology demonstrates outstanding performance in detecting and mitigating insider threats, achieving 100% accuracy based on Decision Tree analysis using CERT and darknet datasets.

CONCLUSION

This research proposed a methodology for detecting and mitigating insider threats using machine learning techniques. In P&ID, Nearmiss-2 based sampling is applied, along with B-SVM algorithm to detect and classify intentional, unintentional insiders, and genuine user. It achieved 95.69% accuracy for detecting 8 intentional and 1 unintentional insider using CERT dataset. It detected 4,783 intentional-darknet and 68 unintentional-darknet users in CIC darknet dataset. In UIM, the CKPCA-DBN is developed to authenticate unintentional insiders and it achieved minimum 0.5% EER and 99.84% accuracy to mitigate one unintentional insider as intentional insider in CERT dataset. In CIC darknet dataset, 68 unintentional-darknet users are mitigated as 64 intentional-darknet and 4 benign users. In IIM, user profiling mechanisms using a decision tree is incorporated to profile 8 intentional insiders into Denylist and 57 geniune users into Allowlist using CERT dataset. In CIC darknet dataset, 5063 benign users are profiled into Allowlist and 4847 intentional-darknet users are profiled into Denylist. This research contributes significantly to security in the field of insider threat.

ACKNOWLEDGMENT

I acknowledge Avinashilingam Institute for Home Science and Higher Education for Women for the support of patent publication. I would also like to acknowledge Center for Cyber Intelligence, and Center for Machine Learning and Intelligence for funding to execute a minimum part of phase-I

REFERENCES

1. RA Alsowail, T Al-Shehari. Techniques and countermeasures for preventing insider threats, PeerJ Computer Science 8 (2022) e938.
2. M Villarreal-Vasquez, GM Howard, S Dube, B Bhargava, Hunting for insider threats using lstm-based anomaly detection, IEEE Transactions on Dependable and Secure Computing 20 (1) (2021) 451-462.
3. W Meng, W Li, Y Xiang, KK Choo, A bayesian inference-based detection mechanism to defend medical smartphone networks against insider attacks, Journal of Network and Computer Applications 78 (2017) 162-169.
4. Forbes, Top Cybersecurity Predictions. https://www.forbes.com/sites/emilsayegh/2022/12/15/top-cybersecurity-predictions-2023/?sh=36224e8c383f , 2023 (accessed 22 Feb 2023).
5. Cyber Security Hub, The biggest data breaches and leaks of 2022. https://www.cshub.com/attacks/articles/the-biggest-data-breaches-and-leaks-of-2022, 2023 (accessed 22 Feb 2023).
6. CrowdStrike, 2022 Global Threat Report. https://irp.cdn-website.com/5d9b1ea1/files/uploaded/Report2022GTR.pdf, 2023 (accessed 22 Feb 2023).

7. Verizon, 2022 Data Breach Investigations Report. https://www.verizon.com/business/resources/T158/reports/dbir/2022-data-breach-investigations-report-dbir.pdf, 2023 (accessed 22 Feb 2023).
8. E Alabdulkreem, M Alduailij, M Alduailij, RF Mansour. Optimal weighted fusion based insider data leakage detection and classification model for Ubiquitous computing systems, Sustainable Energy Technologies and Assessments 54 (2022)102815.
9. W Li, W Meng, LF Kwok, HS Horace, Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model, Journal of Network and Computer Applications 77 (2017) 135-145.
10. D Soni, N Kumar, Machine learning techniques in emerging cloud computing integrated paradigms: A survey and taxonomy, Journal of Network and Computer Applications. 2022 May 24:103419.
11. M Omar, D Mohammed, V Nguyen, Defending against malicious insiders: a conceptual framework for predicting, detecting, and deterring malicious insiders, International Journal of Business Process Integration and Management 8 (2) (2017) 114-9.
12. R Singh, H Kumar, RK Singla, An intrusion detection system using network traffic profiling and online sequential extreme learning machine, Expert Systems with Applications 42 (22) (2015) 8609-8624.
13. S Asha, D Shanmugapriya, G Padmavathi, 2023. Malicious insider threat detection using variation of sampling methods for anomaly detection in cloud environment, Computers and Electrical Engineering 105, 108519.
14. M Canham, C Posey, PS Bockelman, Confronting information security's elephant, the unintentional insider threat, in: Augmented Cognition. Human Cognition and Behavior: 14th International Conference, AC 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22 2020, Springer International Publishing, 2020, pp. 316-334.
15. K.S. Killourhy, and R.A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics", in Proceedings of IEEE/IFIP International Conference on Dependable Systems & Networks, pp. 125-134, 2009.
16. Lindauer, Brian (2020). Insider Threat Test Dataset. Carnegie Mellon University. Dataset. https://doi.org/10.1184/R1/12841247.v1
17. Arash Habibi Lashkari, Gurdip Kaur, and Abir Rahali, **"DIDarknet: A Contemporary Approach to Detect and Characterize the Darknet Traffic using Deep Image Learning"**, 10th International Conference on Communication and Network Security, Tokyo, Japan, November 2020.

PUBLICATIONS AND CONFERENCES

PATENT PUBLISHED

1. SHANMUGAPRIYA, D and ASHA, S. (2024). *A METHODOLOGY FOR DETECTION AND MITIGATION OF INTENTIONAL AND UNINTENTIONAL INSIDERS* (Patent Application No: 202441068274 A). Indian Patent Office.

JOURNAL PUBLICATIONS

1. Asha S, Shanmugapriya D, Understanding insiders in cloud adopted organizations: A survey on taxonomies, incident analysis, defensive solutions, challenges, Future Generation Computer Systems, 158, 2024, pg.no. 427-446, ISSN: 0167-739X, https://doi.org/10.1016/j.future.2024.04.033. **Elsevier (Scopus, SCIE Indexed, UGC CARE LIST-1) Citations: 3 | Impact Factor: 6.2| Cite Score: 19.9**

2. Asha S, Shanmugapriya D, Padmavathi G, Malicious insider threat detection using variation of sampling methods for anomaly detection in cloud environment, Computers and Electrical Engineering, Volume 105, 2023, 108519, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2022.108519. **Elsevier (Scopus, SCIE Indexed, UGC-CARE LIST 1) Citations: 11 | Impact Factor: 4.0 | Cite Score: 9.2**

CONFERENCES

1. G. Padmavathi, D. Shanmugapriya and S. Asha, "A Framework to Detect the Malicious Insider Threat in Cloud Environment using Supervised Learning Methods," *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2022, **IEEE**, pp. 354-358, doi: 10.23919/INDIACom54597.2022.9763205. (**Scopus Indexed**).

2. Padmavathi, G., Shanmugapriya, D., Asha, S. (2022). A Framework for Improving the Accuracy with Different Sampling Techniques for Detection of Malicious Insider Threat in Cloud. In: Uddin, M.S., Jamwal, P.K., Bansal, J.C. (eds) Proceedings of International Joint Conference on Advances in Computational Intelligence. Algorithms for Intelligent Systems. **Springer**, Singapore. https://doi.org/10.1007/978-981-19-0332-8_36

BOOK CHAPTER

1. Padmavathi G, Shanmugapriya D, Asha S. Detection of Malicious Insider in Cloud Environment based on behavior Analysis. In Progressions Made in Cyber-Security World 2022 (pp. 1-9). **CRC Press** (**Taylor & Francis)**